

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- ① • BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup>:</b> <b>C12N 15/55, 15/62, 15/86, 15/65</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 94/28143</b> <b>(43) International Publication Date:</b> 8 December 1994 (08.12.94)
<b>(21) International Application Number:</b> PCT/US94/05601 <b>(22) International Filing Date:</b> 19 May 1994 (19.05.94)  <b>(30) Priority Data:</b> 08/066,388 21 May 1993 (21.05.93) US  <b>(71) Applicant:</b> TARGETED GENETICS CORPORATION [US/US]; Suite 100, 1100 Olive Way, Seattle, WA 98101 (US).  <b>(72) Inventor:</b> LUPTON, Stephen, D.; 3122 N.E. 83rd Street, Seattle, WA 98115 (US).  <b>(74) Agents:</b> DYLAN, Tyler et al.; Morrison & Foerster, 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).		<b>(81) Designated States:</b> AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> BIFUNCTIONAL SELECTABLE FUSION GENES BASED ON THE CYTOSINE DEAMINASE (CD) GENE  <b>(57) Abstract</b> <p>The invention provides selectable fusion genes including a dominant positive selectable gene fused to and in reading frame with a negative selectable gene. The selectable fusion gene encodes a single bifunctional fusion protein which is capable of conferring a dominant positive selectable phenotype and a negative selectable phenotype on a cellular host. A dominant negative selectable phenotype is conferred by the cytosine deaminase (CD) gene for 5-fluorocytosine sensitivity (5-FC<sup>s</sup>). A dominant positive selectable phenotype is conferred, for example, by the <i>neo</i> gene for G-418 aminoglycoside antibiotic resistance (G-418<sup>r</sup>), or by the <i>hph</i> gene for hygromycin B resistance (Hm<sup>r</sup>). The present invention also provides recombinant expression vectors, such as retroviral vectors, which include selectable fusion genes, and cells transduced with the recombinant expression vectors. The bifunctional selectable fusion genes are expressed and regulated as a single genetic entity, permitting co-regulation and co-expression with a high degree of efficiency.</p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

**BIFUNCTIONAL SELECTABLE FUSION GENES**  
**BASED ON THE CYTOSINE DEAMINASE (CD) GENE**

**Background**

5           The present invention relates generally to genes expressing selectable phenotypes. More particularly, the present invention relates to genes capable of co-expressing both dominant positive selectable and negative selectable phenotypes.

          Genes which express a selectable phenotype are widely used in recombinant DNA technology as a means for identifying and isolating host cells into which the gene has been  
10       introduced. Typically, the gene expressing the selectable phenotype is introduced into the host cell as part of a recombinant expression vector. Positive selectable genes provide a means to identify and/or isolate cells that have retained introduced genes in a stable form, and, in this capacity, have greatly facilitated gene transfer and the analysis of gene function. Negative selectable genes, on the other hand, provide a means for eliminating cells that retain the  
15       introduced gene.

          A variety of genes are available which confer selectable phenotypes on animal cells. The bacterial neomycin phosphotransferase (*neo*) (Colbere-Garapin et al., *J. Mol. Biol.* 150:1, 1981), hygromycin phosphotransferase (*hph*) (Santerre et al., *Gene* 30:147, 1984), and xanthine-guanine phosphoribosyl transferase (*gpt*) (Mulligan and Berg, *Proc. Natl. Acad. Sci.*  
20       *USA* 78:2072, 1981) genes are widely used dominant positive selectable genes. The *Herpes simplex* virus type I thymidine kinase (HSV-I TK) gene (Wigler et al., *Cell* 11:223, 1977); the cellular adenine phosphoribosyltransferase (APRT) (Wigler et al., *Proc. Natl. Acad. Sci. USA* 76:1373, 1979); and hypoxanthine phosphoribosyltransferase (HPRT) genes (Jolly et al., *Proc. Natl. Acad. Sci. USA* 80:477, 1983) are commonly used recessive positive selectable genes. In  
25       general, dominant selectable genes are more versatile than recessive genes, because the use of recessive genes is limited to mutant cells deficient in the selectable function, whereas dominant genes may be used in wild-type cells.

          Several genes confer negative as well as positive selectable phenotypes, including the HSV-I TK, HPRT, APRT and *gpt* genes. These genes encode enzymes which catalyze the  
30       conversion of nucleoside or purine analogs to cytotoxic intermediates. The nucleoside analog

ganciclovir (GCV) is an efficient substrate for HSV-I TK, but a poor substrate for cellular TK, and therefore may be used for negative selection against the HSV-I TK gene in wild-type cells (St. Clair et al., *Antimicrob. Agents Chemother.* 31:844, 1987). However, the HSV-I TK gene may only be used effectively for positive selection in mutant cells lacking cellular TK activity.

- 5 Use of the HPRT and APRT genes for either positive or negative selection is similarly limited to HPRT<sup>-</sup> or APRT<sup>-</sup> cells, respectively (Fenwick, "The HGPRT System", pp. 333-373, M. Gottesman (ed.), *Molecular Cell Genetics*, John Wiley and Sons, New York, 1985; Taylor et al., "The APRT System", pp. 311-332, M. Gottesman (ed.), *Molecular Cell Genetics*, John Wiley and Sons, New York, 1985). The *gpt* gene, on the other hand, may be used for both
- 10 positive and negative selection in wild-type cells. Negative selection against the *gpt* gene in wild-type cells is possible using 6-thioxanthine, which is efficiently converted to a cytotoxic nucleotide analog by the bacterial *gpt* enzyme, but not by the cellular HPRT enzyme (Besnard et al., *Mol. Cell. Biol.* 7:4139, 1987).

- Another negatively selectable gene has recently been reported by Mullen et al., *Proc. Natl. Acad. Sci. USA* 89:33, 1992. The bacterial cytosine deaminase (CD) gene converts 5-fluorocytosine (5-FC) to 5-fluorouracil (5-FU). 5-FU is further metabolized intracellularly to 5-fluoro-uridine-5'-triphosphate and 5-fluoro-2'-deoxy-uridine-5'-monophosphate, which inhibit RNA and DNA synthesis, causing cell death. Thus, 5-FC can effectively ablate cells carrying and expressing the CD gene. The CD gene is not positively selectable in normal cells.

- 20 More recently, attention has turned to selectable genes that may be incorporated into gene transfer vectors designed for use in human gene therapy. Gene therapy can be used as a means for augmenting normal cellular function, for example, by introducing a heterologous gene capable of modifying cellular activities or cellular phenotype, or alternatively, expressing a drug needed to treat a disease. Gene therapy may also be used to treat a hereditary genetic
- 25 disease which results from a defect in or absence of one or more genes. Collectively, such diseases result in significant morbidity and mortality. Examples of such genetic diseases include hemophilias A and B (caused by a deficiency of blood coagulation factors VIII and IX, respectively), alpha-1-antitrypsin deficiency, and adenosine deaminase deficiency. In each of these particular cases, the missing gene has been identified and its complementary DNA
- 30 (cDNA) molecularly cloned (Wood et al., *Nature* 312:330, 1984; Anson et al., *Nature* 315:683, 1984; and Long et al., *Biochemistry* 23:4828, 1984; Daddona et al., *J. Biol. Chem.* 259:12101, 1984). While palliative therapy is available for some of these genetic diseases, often in the form of administration of blood products or blood transfusions, one way of treating such genetic diseases is to introduce a replacement for the defective or missing gene back into

the somatic cells of the patient, a process referred to as "gene therapy" (Anderson, *Science* 226:401, 1984).

The process of gene therapy typically involves the steps of (1) removing somatic (non-germ) cells from the patient, (2) introducing into the cells *ex vivo* a therapeutic or replacement gene via an appropriate vector capable of expressing the therapeutic or replacement gene, and (3) transplanting or transfusing these cells back into the patient, where the therapeutic or replacement gene is expressed to provide some therapeutic benefit. Gene transfer into somatic cells for human gene therapy is presently achieved *ex vivo* (Kasid et al., *Proc. Natl. Acad. Sci. USA* 87:473, 1990; Rosenberg et al., *N. Engl. J. Med.* 323:570, 1990), and this relatively inefficient process would be facilitated by the use of a dominant positive selectable gene for identifying and isolating those cells into which the replacement gene has been introduced before they are returned to the patient. The *neo* gene, for example, has been used to identify genetically modified cells used in human gene therapy.

In some instances, however, it is possible that the introduction of genetically modified cells may actually compromise the health of the patient. The ability to selectively eliminate genetically modified cells *in vivo* would provide an additional margin of safety for patients undergoing gene therapy, by permitting reversal of the procedure. This might be accomplished by incorporating into the vector a negative selectable (or "suicide") gene that is capable of functioning in wild-type cells. Incorporation of a gene capable of conferring both dominant positive and negative selectable phenotypes would ensure co-expression and co-regulation of the positive and negative selectable phenotypes, and would minimize the size of the vector. However, positive selection for the *gpt* gene in some instances requires precise selection conditions which may be difficult to determine. For these reasons, co-expression of a dominant positive selectable phenotype and a negative selectable phenotype is typically achieved by co-expressing two different genes which separately encode other dominant positive and negative selectable functions, rather than using the *gpt* gene.

The existing strategies for co-expressing dominant positive and negative selectable phenotypes encoded by different genes often present complex challenges. The most widely used technique is to co-transfect two plasmids separately encoding two phenotypes (Wigler et al., *Cell* 16:777, 1979). However, the efficiency of co-transfer is rarely 100%, and the two genes may be subject to independent genetic or epigenetic regulation. A second strategy is to link the two genes on a single plasmid, or to place two independent transcription units into a viral vector. This method also suffers from the disadvantage that the genes may be independently regulated. In retroviral vectors, suppression of one or the other independent

transcription unit may occur (Emerman and Temin, *Mol. Cell. Biol.* 6:792, 1986). In addition, in some circumstances there may be insufficient space to accommodate two functional transcription units within a viral vector, although retroviral vectors with functional multiple promoters have been successfully made (Overell et al., *Mol. Cell. Biol.* 8:1803, 1988). A third  
5 strategy is to express the two genes as a bicistronic mRNA using a single promoter. With this method, however, the distal open reading frame is often translated with variable (and usually reduced) efficiency (Kaufman et al., *EMBO J.* 6:187, 1987), and it is unclear how effective such an expression strategy would be in primary cells.

The present invention provides a method for more efficiently and reliably co-expressing  
10 a dominant positive selectable phenotype and a negative selectable phenotype encoded by different genes.

#### SUMMARY OF THE INVENTION

The present invention provides a selectable fusion gene comprising a dominant positive  
15 selectable gene fused to and in reading frame with a negative selectable gene. The selectable fusion gene encodes a single bifunctional fusion protein which is capable of conferring a dominant positive selectable phenotype and a negative selectable phenotype on a cellular host. The selectable fusion genes of the present invention comprise nucleotide sequences for negative selection that are derived from the bacterial cytosine deaminase (CD) gene.

20 In a preferred embodiment, the selectable fusion gene comprises nucleotide sequences from the bacterial CD gene fused to nucleotide sequences from the *neo* gene, referred to herein as the CD-*neo* selectable fusion gene (Sequence Listing No. 1). The CD-*neo* selectable fusion gene confers both G-418 resistance ( $G-418^R$ ) for dominant positive selection and 5-fluorocytosine sensitivity ( $5-FC^S$ ) for negative selection.

25 The present invention also provides recombinant expression vectors, for example, retroviruses, which include the selectable fusion genes, and cells transduced with the recombinant expression vectors.

The selectable fusion genes of the present invention are expressed and regulated as a single genetic entity, permitting co-regulation and co-expression with a high degree of  
30 efficiency.

### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows diagrams of the expression cassettes contained in plasmids tgCMV/hygro/LTR, tgCMV/neo, tgCMV/hygro-CD, tgCMV/CD-hygro, tgCMV/neo-CD and tgCMV/CD-neo. The horizontal arrows indicate transcriptional start sites and direction of transcription. The open box labeled LTR is the retroviral long terminal repeat. The open box labeled CMV is the cytomegalovirus promoter.

Figure 2 shows the results of the cytosine deaminase assay on extracts prepared from transfected pools of NIH/3T3 cells. The extracts were assayed by measuring the conversion of cytosine to uracil.

Figure 3 shows diagrams of the proviral structures of retroviral vectors tgLS(+)neo and tgLS(+)CD-neo used in the present invention.

Figure 4 shows the results of the cytosine deaminase assay on uninfected (lane 1), tgLS(+)neo-infected (lane 2) and tgLS(+)CD-neo-infected NIH/3T3 (lane 3) cell pools. The results indicate that cells infected with the tgLS(+)CD-neo express high levels of cytosine deaminase activity.

Figure 5 shows photographs of stained colonies of uninfected NIH/3T3 cells (plates a, b and c) and NIH/3T3 cells infected with the tgLS(+)neo (plates d and e) or tgLS(+)CD-neo (plates f and g) retroviruses. The cells were grown in medium alone (plate a) or medium supplemented with G-418 (plates b, d and f) or G-418+5-FC (plates c, e and g) in a long-term proliferation assay. The data show that uninfected NIH/3T3 cells were sensitive to G-418 and resistant to 5-FC, NIH/3T3 cells infected with tgLS(+)neo are resistant to both G-418 and 5-FC, and NIH/3T3 cells infected with tgLS(+)CD-neo are resistant to G-418 and sensitive to 5-FC.

### DETAILED DESCRIPTION OF THE INVENTION

SEQ ID NO:1 and SEQ ID NO:2 (appearing immediately prior to the claims) show specific embodiments of the nucleotide sequence and corresponding amino acid sequence of the CD-neo selectable fusion gene of the present invention. The CD-neo selectable fusion gene shown in the Sequence Listing comprises sequences from the CD gene (nucleotides 4-1281) linked to sequences from the neo gene (nucleotides 1282-2073).

#### Definitions

As used herein, the term "selectable fusion gene" refers to a nucleotide sequence comprising a dominant positive selectable gene which is fused to and in reading frame with a negative selectable gene and which encodes a single bifunctional fusion protein which is capable



of conferring a dominant positive selectable phenotype and a negative selectable phenotype on a cellular host. A "dominant positive selectable gene" refers to a sequence of nucleotides which encodes a protein conferring a dominant positive selectable phenotype on a cellular host, and is discussed and exemplified in further detail below. A "negative selectable gene" refers to a sequence of nucleotides which encodes a protein conferring a negative selectable phenotype on a cellular host, and is also discussed and exemplified in further detail below. A "selectable gene" refers generically to dominant positive selectable genes and negative selectable genes.

A selectable gene is "fused to and in reading frame with" another selectable gene if the expression products of the selectable genes (i.e., the proteins encoded by the selectable genes) are fused by a peptide bond and at least part of the biological activity of each of the two proteins is retained. With reference to the CD-neo selectable fusion gene disclosed herein, the CD gene (encoding cytosine deaminase, which confers a negative selectable phenotype of 5-fluorocytosine sensitivity, or 5-FC<sup>S</sup>) is fused to and in reading frame with the neo gene (encoding neomycin phosphotransferase, which confers the dominant positive selectable phenotype of G-418 resistance, or G-418<sup>r</sup>) if the CD and neo proteins are fused by a peptide bond and expressed as a single bifunctional fusion protein.

The component selectable gene sequences of the present invention are preferably contiguous; however, it is possible to construct selectable fusion genes in which the component selectable gene sequences are separated by internal nontranslated nucleotide sequences, such as introns. For purposes of the present invention, such noncontiguous selectable gene sequences are considered to be fused, provided that expression of the selectable fusion gene results in a single bifunctional fusion protein in which the expression products of the component selectable gene sequences are fused by a peptide bond.

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides or ribonucleotides, such as a DNA or RNA sequence. Nucleotide sequences may be in the form of a separate fragment or as a component of a larger construct. Preferably, the nucleotide sequences are in a quantity or concentration enabling identification, manipulation, and recovery of the sequence by standard biochemical methods, for example, using a cloning vector. Recombinant nucleotide sequences are the product of various combinations of cloning, restriction, and ligation steps resulting in a construct having a structural coding sequence distinguishable from homologous sequences found in natural systems. Generally, nucleotide sequences encoding the structural coding sequence, for example, the selectable fusion genes of the present invention, can be assembled from nucleotide fragments and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of

being expressed in a recombinant transcriptional unit. Such sequences are preferably provided in the form of an open reading frame uninterrupted by internal nontranslated sequences, or introns, which are typically present in eukaryotic genes. Genomic DNA containing the relevant selectable gene sequences is preferably used to obtain appropriate nucleotide sequences  
5 encoding selectable genes; however, cDNA fragments may also be used. Sequences of non-translated DNA may be present 5' or 3' from the open reading frame or within the open reading frame, provided such sequences do not interfere with manipulation or expression of the coding regions. Some genes, however, may include introns which are necessary for proper expression in certain hosts, for example, the HPRT selectable gene includes introns which are  
10 necessary for expression in embryonal stem (ES) cells. As suggested above, the nucleotide sequences of the present invention may also comprise RNA sequences, for example, where the nucleotide sequences are packaged as RNA in a retrovirus for infecting a cellular host. The use of retroviral expression vectors is discussed in greater detail below.

The term "recombinant expression vector" refers to a replicable unit of DNA or RNA  
15 in a form which is capable of being transduced into a target cell by transfection or viral infection, and which codes for the expression of a selectable fusion gene which is transcribed into mRNA and translated into protein under the control of a genetic element or elements having a regulatory role in gene expression, such as transcription and translation initiation and termination sequences. The recombinant expression vectors of the present invention can take  
20 the form of DNA constructs replicated in bacterial cells and transfected into target cells directly, for example, by calcium phosphate precipitation, electroporation or other physical transfer methods. The recombinant expression vectors which take the form of RNA constructs may, for example, be in the form of infectious retroviruses packaged by suitable "packaging" cell lines which have previously been transfected with a proviral DNA vector and produce a  
25 retrovirus containing an RNA transcript of the proviral DNA. A host cell is infected with the retrovirus, and the retroviral RNA is replicated by reverse transcription into a double-stranded DNA intermediate which is stably integrated into chromosomal DNA of the host cell to form a provirus. The provirus DNA is then expressed in the host cell to produce polypeptides encoded by the DNA. The recombinant expression vectors of the present invention thus include not  
30 only RNA constructs present in the infectious retrovirus, but also copies of proviral DNA, which include DNA reverse transcripts of a retrovirus RNA genome stably integrated into chromosomal DNA in a suitable host cell, or cloned copies thereof, or cloned copies of unintegrated intermediate forms of retroviral DNA. Proviral DNA includes transcriptional elements in independent operative association with selected structural DNA sequences which are  
35 transcribed into mRNA and translated into protein when proviral sequences are expressed in

infected host cells. Recombinant expression vectors used for direct transfection will include DNA sequences enabling replication of the vector in bacterial host cells. Various recombinant expression vectors suitable for use in the present invention are described below.

"Transduce" means introduction of a recombinant expression vector containing a  
5 selectable fusion gene into a cell. Transduction methods may be physical in nature (i.e.,  
transfection), or they may rely on the use of recombinant viral vectors, such as retroviruses,  
encoding DNA which can be transcribed to RNA, packaged into infectious viral particles and  
used to infect target cells and thereby deliver the desired genetic material (i.e., infection).  
Many different types of mammalian gene transfer and recombinant expression vectors have  
10 been developed (see, e.g., Miller and Calos, Eds., "Gene Transfer Vectors for Mammalian  
Cells," *Current Comm. Mol. Biol.*, (Cold Spring Harbor Laboratory, New York, 1987)).  
Naked DNA can be physically introduced into mammalian cells by transfection using any one  
of a number of techniques including, but not limited to, calcium phosphate transfection (Berman  
et al., *Proc. Natl. Acad. Sci. USA* 84 81:7176, 1984), DEAE-Dextran transfection (McCutchan  
15 et al., *J. Natl. Cancer Inst.* 41:351, 1986; Luthman et al., *Nucl. Acids Res.* 11:1295, 1983),  
protoplast fusion (Deans et al., *Proc. Natl. Acad. Sci. USA* 84 81:1292, 1984), electroporation  
(Potter et al., *Proc. Natl. Acad. Sci. USA* 84 81:7161, 1984), lipofection (Felgner et al., *Proc.*  
*Natl. Acad. Sci. USA* 84:7413, 1987), Polybrene hexadimethrine bromide transfection (Kawai  
and Nishizawa, *Mol. Cell. Biol.* 4:1172, 1984) and direct gene transfer by laser micropuncture  
20 of cell membranes (Tao et al., *Proc. Natl. Acad. Sci. USA* 84:4180, 1987). Various infection  
techniques have been developed which utilize recombinant infectious virus particles for gene  
delivery. This represents a preferred approach to the present invention. The viral vectors  
which have been used in this way include virus vectors derived from simian virus 40 (SV40;  
Karlsson et al., *Proc. Natl. Acad. Sci. USA* 84 82:158, 1985), adenoviruses (Karlsson et al.,  
25 *EMBO J.* 5:2377, 1986), adeno-associated virus (LaFace et al., *Virology* 162:483, 1988) and  
retroviruses (Coffin, 1985, p17-71 in Weiss et al. (eds.), *RNA Tumor Viruses*, 2nd ed. Vol 2,  
Cold Spring Harbor Laboratory, New York). Thus, gene transfer and expression methods are  
numerous but essentially function to introduce and express genetic material in mammalian cells.  
Several of the above techniques have been used to transduce hematopoietic or lymphoid cells,  
30 including calcium phosphate transfection (Berman et al., *supra*, 1984), protoplast fusion (Deans  
et al., *supra*, 1984), electroporation (Cann et al., *Oncogene* 3:123, 1988), and infection with  
recombinant adenovirus (Karlsson et al., *supra*; Reuther et al., *Mol. Cell. Biol.* 6:123, 1986)  
adeno-associated virus (LaFace et al., *supra*) and retrovirus vectors (Overell et al., *Oncogene*  
4:1425, 1989). Primary T lymphocytes have been successfully transduced by electroporation

(Cann et al., *supra*, 1988) and by retroviral infection (Nishihara et al., *Cancer Res.* 48:4730, 1988; Kasid et al., *supra*, 1990).

#### Construction of Selectable Fusion Genes

5           The selectable fusion genes of the present invention comprise a dominant positive selectable gene fused to a negative selectable gene. A selectable gene will generally comprise, for example, a gene encoding a protein capable of conferring an antibiotic resistance phenotype or supplying an autotrophic requirement (for dominant positive selection), or activating a toxic metabolite (for negative selection). A DNA sequence encoding a bifunctional fusion protein is  
10       constructed using recombinant DNA techniques to assemble separate DNA fragments encoding a dominant positive selective gene and a negative selectable gene into an appropriate expression vector. The 3' end of the one selectable gene is ligated to the 5' end of the other selectable gene, with the reading frames of the sequences in frame to permit translation of the mRNA sequences into a single biologically active bifunctional fusion protein. The selectable fusion  
15       gene is expressed under control of a single promoter.

          The dominant positive selectable gene is a gene which, upon being transduced into a host cell, expresses a dominant phenotype permitting positive selection of stable transductants. The dominant positive selectable gene of the present invention is preferably selected from the group consisting of the aminoglycoside phosphotransferase gene (*neo* or *aph*) from Tn5 which  
20       codes for resistance to the antibiotic G418 (Colbere-Garapin et al., *J. Mol. Biol.* 150:1, 1981; Southern and Berg, *J. Mol. Appl. Genet.* 1:327, 1982); and the hygromycin-B phosphotransferase gene (*hph* or "hygro") which confers the selectable phenotype of hygromycin resistance ( $Hm^r$ ) (Santerre et al., *Gene* 30:147, 1984; Sugden et al., *Mol. Cell. Biol.* 5:410, 1985; obtainable from plasmid pHEBol, under ATCC Accession No. 39820).  
25       Hygromycin B is an aminoglycoside antibiotic that inhibits protein synthesis by disrupting translocation and promoting mistranslation. The *hph* gene confers  $Hm^r$  to cells transduced with the *hph* gene by phosphorylating and detoxifying the antibiotic hygromycin B. Other acceptable dominant positive selectable genes include the following: the bacterial *neo* gene encoding neomycin phosphotransferase (Beck et al., *Gene* 19:327, 1982); the xanthine-guanine  
30       phosphoribosyl transferase gene (*gpt*) from *E. coli* encoding resistance to mycophenolic acid (Mulligan and Berg, *Proc. Natl. Acad. Sci. USA* 78:2072, 1981); the dihydrofolate reductase (DHFR) gene from murine cells or *E. coli* which is necessary for biosynthesis of purines and can be competitively inhibited by the drug methotrexate (MTX) to select for cells constitutively expressing increased levels of DHFR (Simonsen and Levinson, *Proc. Natl. Acad. Sci. USA*

80:2495, 1983; Simonsen et al., *Nucl. Acids Res.* 16:2235, 1988); the *S. typhimurium* histidinol dehydrogenase (*hisD*) gene (Hartman et al., *Proc. Natl. Acad. Sci. USA* 85:8047, 1988); the *E. coli* tryptophan synthase  $\beta$  subunit (*trpB*) gene (Hartman et al., *supra*); the puromycin-N-acetyl transferase (*pac*) gene (Vara et al., *Nucl. Acids Res.* 14:4117, 1986); the adenosine deaminase (ADA) gene (Daddona et al., *J. Biol. Chem.* 259:12101, 1984); the multi-drug resistance (MDR) gene (Kane et al., *Gene* 84:439, 1989); the mouse ornithine decarboxylase (OCD) gene (Gupba and Coffino, *J. Biol. Chem.* 160:2941, 1985); the *E. coli* aspartate transcarbamylase catalytic subunit (*pyrB*) gene (Ruiz and Wahl, *Mol. Cell. Biol.* 6:3050, 1986); and the *E. coli* *asnA* gene, encoding asparagine synthetase (Cartier et al., *Mol. Cell. Biol.* 7:1623, 1987).

10       The negative selectable gene is a gene which, upon being transduced into a host cell, expresses a phenotype permitting negative selection (i.e., elimination) of stable transductants. The preferred negative selectable gene of the present invention is the bacterial CD gene encoding cytosine deaminase (Genbank accession number X63656) which confers 5-fluorocytosine sensitivity.

15       Other enzymes suitable for negative selection include, but are not limited to, alkaline phosphatase useful for converting phosphate-containing prodrugs such as etoposide-phosphate, doxorubicin-phosphate, mitomycin phosphate, into toxic dephosphorylated metabolites; arylsulfatase useful for converting sulfate-containing prodrugs into free drugs; proteases, such as serratia protease, thermolysin, subtilisin, carboxypeptidases and cathepsins (such as  
20       cathepsins B and L), that are useful for converting peptide-containing prodrugs into free drugs; D-alanylcarboxypeptidases, useful for converting prodrugs that contain D-amino acid substituents; carbohydrate-cleaving enzymes such as  $\beta$ -galactosidase and neuraminidase useful for converting glycosylated prodrugs into free drugs;  $\beta$ -lactamase useful for converting drugs derivatized with  $\beta$ -lactams into free drugs; and penicillin amidases, such as penicillin V amidase  
25       or penicillin G amidase, useful for converting drugs derivatized at their amino nitrogens with phenoxyacetyl or phenylacetyl groups, respectively, into free drugs.

Other enzyme prodrug combinations include the bacterial (for example, from *Pseudomonas*) enzyme carboxypeptidase G2 with the prodrug para-N-bis(2-chloroethyl) aminobenzoyl glutamic acid. Cleavage of the glutamic acid moiety from this compound  
30       releases a toxic benzoic acid mustard. Penicillin-V amidase will convert phenoxyacetamide derivatives of doxorubicin and melphalan to toxic metabolites.

Due to the degeneracy of the genetic code, there can be considerable variation in nucleotide sequences encoding the same amino acid sequence; exemplary DNA embodiments are those corresponding to the nucleotide sequences in Sequence Listing No. 1. Such variants

will have modified DNA or amino acid sequences, having one or more substitutions, deletions, or additions, the net effect of which is to retain biological activity, and may be substituted for the specific sequences disclosed herein. The sequences of selectable fusion genes comprising CD and *neo* are equivalent if they contain all or part of the sequences of CD and *neo* and are capable of hybridizing to the nucleotide sequence of Sequence Listing No. 1 under moderately stringent conditions (50°C, 2 X SSC) and express a biologically active fusion protein.

A "biologically active" fusion protein will share sufficient amino acid sequence similarity with the specific embodiments of the present invention disclosed herein to be capable of conferring the selectable phenotypes of the component selectable genes.

10 In a preferred embodiment, sequences from the bacterial cytosine deaminase (CD) gene are fused with sequences from the bacterial neomycin phosphotransferase (*neo*) gene. The resulting selectable fusion gene (referred to as the CD-*neo* selectable fusion gene) encodes a bifunctional fusion protein that confers G-418<sup>r</sup> and 5-GC<sup>s</sup> and provides a means by which dominant positive and negative selectable phenotypes may be expressed and regulated as a single genetic entity. The CD-*neo* selectable fusion gene may be especially advantageous in patient populations likely to receive ganciclovir.

#### Recombinant Expression Vectors

20 The selectable fusion genes of the present invention are utilized to identify, isolate or eliminate host cells into which the selectable fusion genes are introduced. The selectable fusion genes are introduced into the host cell by transducing into the host cell a recombinant expression vector which contains the selectable fusion gene. Such host cells include cell types from higher eukaryotic origin, such as mammalian or insect cells, or cell types from lower prokaryotic origin.

25 As indicated above, such selectable fusion genes are preferably introduced into a particular cell as a component of a recombinant expression vector which is capable of expressing the selectable fusion gene within the cell and conferring a selectable phenotype. Such recombinant expression vectors generally include synthetic or natural nucleotide sequences comprising the selectable fusion gene operably linked to suitable transcriptional or translational control sequences, for example, an origin of replication, optional operator sequences to control transcription, a suitable promoter and enhancer linked to the gene to be expressed, and other 5' or 3' flanking nontranscribed sequences, and 5' or 3' nontranslated sequences, such as necessary ribosome binding sites, a polyadenylation site, splice donor and acceptor sites, and transcriptional termination sequences. Such regulatory sequences can be derived from

mammalian, viral, microbial or insect genes. Nucleotide sequences are operably linked when they are functionally related to each other. For example, a promoter is operably linked to a selectable fusion gene if it controls the transcription of the selectable fusion gene; or a ribosome binding site is operably linked to a selectable fusion gene if it is positioned so as to permit  
5 translation of the selectable fusion gene into a single bifunctional fusion protein. Generally, operably linked means contiguous.

Specific recombinant expression vectors for use with mammalian, bacterial, and yeast cellular hosts are described by Pouwels et al. (*Cloning Vectors: A Laboratory Manual*, Elsevier, New York, 1985) and are well-known in the art. A detailed description of  
10 recombinant expression vectors for use in animal cells can be found in Rigby, *J. Gen. Virol.* 64:255, 1983; Elder et al., *Ann. Rev. Genet.* 15:295, 1981; and Subramani et al., *Anal. Biochem.* 135:1, 1983. Appropriate recombinant expression vectors may also include viral vectors, in particular retroviruses (discussed in detail below).

The selectable fusion genes of the present invention are preferably placed under the  
15 transcriptional control of a strong enhancer and promoter expression cassette. Examples of such expression cassettes include the human cytomegalovirus immediate-early (HCMV-IE) promoter (Boshart et al., *Cell* 41:521, 1985), the  $\beta$ -actin promoter (Gunning et al., *Proc. Natl. Acad. Sci. USA* 84:5831, 1987), the histone H4 promoter (Guild et al., *J. Virol.* 62:3795, 1988), the mouse metallothionein promoter (McIvor et al., *Mol. Cell. Biol.* 7:838, 1987), the  
20 rat growth hormone promoter (Miller et al., *Mol. Cell Biol.* 5:431, 1985), the human adenosine deaminase promoter (Hantzopoulos et al., *Proc. Natl. Acad. Sci. USA* 86:3519, 1989) the HSV TK promoter (Tabin et al., *Mol. Cell. Biol.* 2:426, 1982), the  $\alpha$ -1 antitrypsin enhancer (Peng et al., *Proc. Natl. Acad. Sci. USA* 85:8146, 1988) and the immunoglobulin enhancer/promoter (Blankenstein, et al., *Nucleic Acid Res.* 16:10939, 1988), the SV40 early or late promoters, the  
25 Adenovirus 2 major late promoter, or other viral promoters derived from polyoma virus, bovine papilloma virus, or other retroviruses or adenoviruses. The promoter and enhancer elements of immunoglobulin (Ig) genes confer marked specificity to B lymphocytes (Banerji et al., *Cell* 33:729, 1983; Gillies et al., *Cell* 33:717, 1983; Mason et al., *Cell* 41:479, 1985), while the elements controlling transcription of the  $\beta$ -globin gene function only in erythroid cells  
30 (van Assendelft et al., *Cell* 56:969, 1989). Using well-known restriction and ligation techniques, appropriate transcriptional control sequences can be excised from various DNA sources and integrated in operative relationship with the intact selectable fusion genes to be expressed in accordance with the present invention. Thus, many transcriptional control

sequences may be used successfully in retroviral vectors to direct the expression of inserted genes in infected cells.

### Retroviruses

- 5           Retroviruses can be used for highly efficient transduction of the selectable fusion genes of the present invention into eukaryotic cells and are preferred for the delivery of a selectable fusion gene into primary cells. Moreover, retroviral integration takes place in a controlled fashion and results in the stable integration of one or a few copies of the new genetic information per cell.
- 10           Retroviruses are a class of viruses whose genome is in the form of RNA. The genomic RNA of a retrovirus contains *trans*-acting gene sequences coding for viral proteins, including: structural proteins (encoded by the *gag* region) that associate with the RNA in the core of the virus particle; reverse transcriptase (encoded by the *pol* region) that makes the DNA complement; and an envelope glycoprotein (encoded by the *env* region) that resides in the
- 15 lipoprotein envelope of the particles and binds the virus to the surface of host cells on infection. Replication of the retrovirus is regulated by *cis*-acting elements, such as the promoter for transcription of the proviral DNA and other nucleotide sequences necessary for viral replication. The *cis*-acting elements are present in or adjacent to two identical untranslated long terminal repeats (LTRs) of about 600 base pairs present at the 5' and 3' ends of the retroviral
- 20 genome. Retroviruses replicate by copying their RNA genome by reverse transcription into a double-stranded DNA intermediate, using a virus-encoded, RNA-directed DNA polymerase, or reverse transcriptase. The DNA intermediate is integrated into chromosomal DNA of an avian or mammalian host cell. The integrated retroviral DNA is called a provirus. The provirus serves as template for the synthesis of RNA chains for the formation of infectious virus
- 25 particles. Forward transcription of the provirus and assembly into infectious virus particles occurs in the presence of an appropriate helper virus having endogenous *trans*-acting genes required for viral replication.

- Retroviruses are used as vectors by replacing one or more of the endogenous *trans*-acting genes of a proviral form of the retrovirus with a recombinant therapeutic gene or, in the
- 30 case of the present invention, a selectable fusion gene, and then transducing the recombinant provirus into a cell. The *trans*-acting genes include the *gag*, *pol* and *env* genes which encode, respectively, proteins of the viral core, the enzyme reverse transcriptase and constituents of the envelope protein, all of which are necessary for production of intact virions. Recombinant retroviruses deficient in the *trans*-acting *gag*, *pol* or *env* genes cannot synthesize essential



proteins for replication and are accordingly replication-defective. Such replication-defective recombinant retroviruses are propagated using packaging cell lines. These packaging cell lines contain integrated retroviral genomes which provide all *trans*-acting gene sequences necessary for production of intact virions. Proviral DNA sequences which are transduced into such

5 packaging cells lines are transcribed into RNA and encapsidated into infectious virions containing the selectable fusion gene (and/or therapeutic gene), but, lacking the *trans*-acting gene products *gag*, *pol* and *env*, cannot synthesize the necessary *gag*, *pol* and *env* proteins for encapsidating the RNA into particles for infecting other cells. The resulting infectious retrovirus vectors can therefore infect other cells and integrate a selectable fusion gene into the

10 cellular DNA of a host cell, but cannot replicate. Mann et al. (*Cell* 33:153, 1983), for example, describe the development of various packaging cell lines (e.g., Ψ2) which can be used to produce helper virus-free stocks of recombinant retrovirus. Encapsidation in a cell line harboring *trans*-acting elements encoding an ecotropic viral envelope (e.g., Ψ2) provides ecotropic (limited host range) progeny virus. Alternatively, assembly in a cell line containing

15 amphotropic packaging genes (e.g., PA317, ATCC CRL 9078; Miller and Buttimore, *Mol. Cell. Biol.* 6:2895, 1986) provides amphotropic (broad host range) progeny virus.

Numerous provirus constructs have been used successfully to express foreign genes (see, e.g., Coffin, in Weiss et al. (eds.), *RNA Tumor Viruses*, 2nd Ed., Vol. 2, (Cold Spring Harbor Laboratory, New York, 1985, pp. 17-71). Most proviral elements are derived from

20 murine retroviruses. Retroviruses adaptable for use in accordance with the present invention can, however, be derived from any avian or mammalian cell source. Suitable retroviruses must be capable of infecting cells which are to be the recipients of the new genetic material to be transduced using the retroviral vector. Examples of suitable retroviruses include avian retroviruses, such as avian erythroblastosis virus (AEV), avian leukemia virus (ALV), avian

25 myeloblastosis virus (AMV), avian sarcoma virus (ASV), Fujinami sarcoma virus (FuSV), spleen necrosis virus (SNV), and Rous sarcoma virus (RSV); bovine leukemia virus (BLV); feline retroviruses, such as feline leukemia virus (FeLV) or feline sarcoma virus (FeSV); murine retroviruses, such as murine leukemia virus (MuLV); mouse mammary tumor virus (MMTV), and murine sarcoma virus (MSV); and primate retroviruses, such as human T-cell

30 lymphotropic viruses 1 and 2 (HTLV-1, and -2), and simian sarcoma virus (SSV). Many other suitable retroviruses are known to those skilled in the art. A taxonomy of retroviruses is provided by Teich, in Weiss et al. (eds.), *RNA Tumor Viruses*, 2d ed., Vol. 2 (Cold Spring Harbor Laboratory, New York, 1985, pp. 1-160). Preferred retroviruses for use in connection with the present invention are the murine retroviruses known as Moloney murine leukemia

virus (MoMLV), Moloney murine sarcoma virus (MoMSV), Harvey murine sarcoma virus (HaMSV) and Kirsten murine sarcoma virus (KiSV). The sequences required to construct a retroviral vector from the MoMSV genome can be obtained in conjunction with a pBR322 plasmid sequence such as pMV (ATCC 37190), while a cell line producer of KiSV in K-BALB  
5 cells has been deposited as ATCC CCL 163.3. A deposit of pRSVneo, derived from pBR322 including the RSV LTR and an intact neomycin drug resistance marker is available from ATCC under Accession No. 37198. Plasmid pPB101 comprising the SNV genome is available as ATCC 45012. The viral genomes of the above retroviruses are used to construct replication-defective retrovirus vectors which are capable of integrating their viral genomes into the  
10 chromosomal DNA of an infected host cell but which, once integrated, are incapable of replication to provide infectious virus, unless the cell in which it is introduced contains other proviral elements encoding functional active *trans*-acting viral proteins.

The selectable fusion genes of the present invention which are transduced by retroviruses are expressed by placing the selectable fusion gene under the transcriptional control  
15 of the enhancer and promoter incorporated into the retroviral LTR, or by placing them under the control of heterologous transcriptional control sequences inserted between the LTRs. Use of both heterologous transcriptional control sequences and the LTR transcriptional control sequences enables coexpression of a therapeutic gene and a selectable fusion gene in the vector, thus allowing selection of cells expressing specific vector sequences encoding the desired  
20 therapeutic gene product. Obtaining high-level expression may require placing the therapeutic gene and/or selectable fusion gene within the retrovirus under the transcriptional control of a strong heterologous enhancer and promoter expression cassette. Many different heterologous enhancers and promoters have been used to express genes in retroviral vectors. Such enhancers or promoters can be derived from viral or cellular sources, including mammalian genomes, and  
25 are preferably constitutive in nature. Such heterologous transcriptional control sequences are discussed above with reference to recombinant expression vectors. To be expressed in the transduced cell, DNA sequences introduced by any of the above gene transfer methods are usually expressed under the control of an RNA polymerase II promoter.

Particularly preferred recombinant expression vectors include pLXSN, pLNCX and  
30 pLNL6, and derivatives thereof, which are described by Miller and Rosman, *Biotechniques* 7:980, 1989. These vectors are capable of expressing heterologous DNA under the transcriptional control of the retroviral LTR or the CMV promoter, and the *neo* gene under the control of the SV40 early region promoter or the retroviral LTR. For use in the present invention, the *neo* gene is replaced with the bifunctional selectable fusion genes disclosed

herein, such as the CD-neo selectable fusion gene. Construction of useful replication-defective retroviruses is a matter of routine skill. The resulting recombinant retroviruses are capable of integration into the chromosomal DNA of an infected host cell, but once integrated, are incapable of replication to provide infectious virus, unless the cell in which it is introduced contains another proviral insert encoding functionally active *trans*-acting viral proteins.

#### Uses of Bifunctional Selectable Fusion Genes

The selectable fusion genes of the present invention are particularly preferred for use in gene therapy as a means for identifying, isolating or eliminating cells, such as somatic cells, into which the selectable fusion genes are introduced. In gene therapy, somatic cells are removed from a patient, transduced with a recombinant expression vector containing a therapeutic gene and the selectable fusion gene of the present invention, and then reintroduced back into the patient. Somatic cells which can be used as vehicles for gene therapy include hematopoietic (bone marrow-derived) cells, keratinocytes, hepatocytes, endothelial cells and fibroblasts (Friedman, *Science* 244:1275, 1989). Alternatively, gene therapy can be accomplished through the use of injectable vectors which transduce somatic cells *in vivo*. The feasibility of gene transfer in humans has been demonstrated (Kasid et al., *Proc. Natl. Acad. Sci. USA* 87:473, 1990; Rosenberg et al., *N. Engl. J. Med.* 323:570, 1990).

The selectable fusion genes of the present invention are particularly useful for eliminating genetically modified cells *in vivo*. *In vivo* elimination of cells expressing a negative selectable phenotype is particularly useful in gene therapy as a means for ablating a cell graft, thereby providing a means for reversing the gene therapy procedure. For example, it has been shown that administration of the anti-herpes virus drug ganciclovir to transgenic animals expressing the HSV-I TK gene from an immunoglobulin promoter results in the selective ablation of cells expressing the HSV-I TK gene (Heyman et al., *Proc. Natl. Acad. Sci. USA* 86:2698, 1989). Using the same transgenic mice, GCV has also been shown to induce full regression of Abelson leukemia virus-induced lymphomas (Moolten et al., *Human Gene Therapy* 1:125, 1990). In a third study, in which a murine sarcoma (K3T3) was infected with a retrovirus expressing HSV-I TK and transplanted into syngeneic mice, the tumors induced by the sarcoma cells were completely eradicated following treatment with GCV (Moolten and Wells, *J. Natl. Cancer Inst.* 82:297, 1990).

The selectable fusion genes of the present invention also are beneficial in tumor ablation therapy as it has been practiced by Oldfield et al., *Human Gene Therapy* 4:39, 1993. Packaging cells (about  $10^6$  -  $10^9$ ) producing the tgLS(+)CD-neo retroviral vectors are

inoculated intra-tumorally. After a period of several days, during which the newly produced retroviruses infect the adjacent rapidly growing tumor cells, the patient is given about 50-200 mg of 5-FC per kg body weight (orally or intravenously) daily (when the tgLS(+)CD-neo retroviral vector has been used) to selectively ablate the infected tumor cells.

5           The bifunctional selectable fusion genes of the present invention can also be used to facilitate gene modification by homologous recombination. Reid et al., *Proc. Natl. Acad. Sci. USA* 87:4299, 1990 has recently described a two-step procedure for gene modification by homologous recombination in ES cells ("in-out" homologous recombination) using the HPRT gene. Briefly, this procedure involves two steps: an "in" step in which the HPRT gene is  
10       embedded in target gene sequences, transfected into HPRT<sup>-</sup> host cells and homologous recombinants having incorporated the HPRT gene into the target locus are identified by their growth in HAT medium and genomic analysis using PCR. In a second "out" step, a construct containing the desired replacement sequences embedded in the target gene sequences (but without the HPRT gene) is transfected into the cells and homologous recombinants having the  
15       replacement sequences (but not the HPRT gene) are isolated by negative selection against HPRT<sup>+</sup> cells. Although this procedure allows the introduction of subtle mutations into a target gene without introducing selectable gene sequences into the target gene, it requires positive selection of transformants in a HPRT<sup>-</sup> cell line, since the HPRT gene is recessive for positive selection. Also, due to the inefficient expression of the HPRT gene in ES cells, it is necessary  
20       to use a large 9-kbp HPRT mini-gene which complicates the construction and propagation of homologous recombination vectors. The selectable fusion genes of the present invention provide an improved means whereby "in-out" homologous recombination may be performed. Because the selectable fusion genes of the present invention are dominant for positive selection, any wild-type cell may be used (i.e., one is not limited to use of cells deficient in the selectable  
25       phenotype). Moreover, the size of the vector containing the selectable fusion gene is reduced significantly relative to the large HPRT mini-gene.

By way of illustration, the CD-neo selectable fusion gene is used as follows: In the first "in" step, the CD-neo selectable fusion gene is embedded in target gene sequences, transfected into a host cell, and homologous recombinants having incorporated the CD-neo  
30       selectable fusion gene into the target locus are identified by their growth in medium containing G-418 followed by genome analysis using PCR. The CD-neo<sup>+</sup> cells are then used in the second "out" step, in which a construct containing the desired replacement sequences embedded in the target gene sequences (but without the CD-neo selectable fusion gene) is transfected into

the cells. Homologous recombinants are isolated by selective elimination of CD-neo<sup>+</sup> cells using 5-FC followed by genome analysis using PCR.

## EXAMPLES

5

### Example 1

#### Construction and Characterization of Plasmid Vectors Containing CD-neo Selectable Fusion Gene

##### 10 A. Construction of the Bifunctional CD-neo Selectable Fusion Gene.

Plasmid tgCMV/hygro/LTR (Figure 1) consists of the following elements: the BalI-SstII fragment containing the HCMV IE94 promoter (Boshart et al., *Cell* 41:521, 1985); an oligonucleotide containing a sequence conforming to a consensus translation initiation sequence for mammalian cells (GCCGCCACC ATG) (Kozak et al., *Nucl. Acids Res.* 15:8125, 1987);  
15 nucleotides 234-1256 from the *hph* gene (Kaster et al., *Nucl. Acids Res.* 11:6895, 1983), encoding hygromycin phosphotransferase; sequences from nucleotide 7764 and through the 3' LTR of MoMLV (Shinnick et al., *Nature* 293:543, 1981), containing a polyadenylation sequence; the NruI-AlwNI fragment from pML2d (Lusky and Botchan, *Nature* 293:79, 1981), containing the bacterial replication origin; the AlwNI-AatII fragment from pGEM1 (Promega Corp.), containing the  $\beta$ -lactamase gene.  
20

Plasmids tgCMV/neo, tgCMV/CD, tgCMV/CD-hygro, tgCMV/neo-CD, and tgCMV/CD-neo are all similar in structure to tgCMV/hygro/LTR and contain the consensus translation initiation sequence; however, each contains different sequences in place of the *hph* sequences. Plasmid tgCMV/neo contains an oligonucleotide encoding three amino acids (GGA TCG GCC) and nucleotide 154-945 from the bacterial *neo* gene encoding neomycin phosphotransferase (Beck et al., *Gene* 19:327, 1982), in place of the *hph* sequences. Plasmid  
25 tgCMV/CD contains nucleotides 1645-2925 from the bacterial CD gene encoding cytosine deaminase (Genbank accession number X63656), in place of the *hph* sequences. The CD sequences were amplified by PCR from plasmid pCD2 (Mullen et al., *Proc. Natl. Acad. Sci. USA* 89:33, 1992). Plasmid tgCMV/hygro-CD contains nucleotides 234-1205 from the *hph* gene fused to nucleotides 1645-2925 from the CD gene in place of the *hph* sequences. Plasmid  
30 tgCMV/CD-hygro contains nucleotides 1645-2922 from the CD gene fused to nucleotides 234-1256 from the *hph* gene in place of the *hph* sequences. Plasmid tgCMV/neo-CD contains an oligonucleotide encoding an additional three amino acids (GGA TCG GCC) and nucleotides

154-942 from the bacterial *neo* gene fused to nucleotides 1645-2925 from the CD gene in place of the *hph* sequences. Plasmid tgCMV/CD-neo contains nucleotides 1645-2922 from the CD gene fused to nucleotides 154-945 from the *neo* gene in place of the *hph* sequences.

Plasmid tgCMV/hygro/LTR was constructed using standard techniques (Ausubel et al.,  
5 *Current Protocols in Molecular Biology* (Wiley, New York), 1987) as follows: Plasmid HyTK-CMV-IL2 was constructed first by ligating the large HindIII-StuI fragment from tgLS(+)HyTK (Lupton et al., *Mol. Cell. Biol.* 11:3374, 1991) with the HindIII-StuI fragment spanning the HCMV IE94 promoter from tgLS(-)CMV/HyTK (Lupton et al., *supra*, 1991), and a fragment containing human IL-2 cDNA sequences. The fragment containing human IL-2 cDNA  
10 sequences was amplified from a plasmid containing the human IL-2 cDNA by PCR using oligonucleotides

5'-CCCGCTAGCCGCCACCATGTACAGGATGCAACTCC-3' and

5'-CCCGTCGACTTAATTATCAAGTCAGTGTT-3'. Following amplification, the PCR product was first treated with T4 DNA polymerase to render the ends blunt, then digested with  
15 NheI, before ligation to the fragments from tgLS(+)HyTK and tgLS(-)CMV/HyTK. To generate plasmid tgCMV/hygro/LTR, the Sall-PvuI fragment spanning the SV40 polyadenylation signal of tgCMV/hygro (Lupton et al., *supra*, 1991) was replaced with the Sall-PvuI fragment containing the Moloney leukemia virus LTR (which contains the retroviral polyadenylation signal) from HyTK-CMV-IL2.

20 Plasmid tgCMV/neo was constructed using standard techniques (Ausubel et al., *supra*, 1987) as follows: A PvuI-NheI fragment spanning the HCMV IE94 promoter from tgCMV/hygro was ligated to a NheI-HindIII fragment spanning the *neo* gene from tgLS(+)neo (the HindIII site was treated with T4 DNA polymerase to render the end blunt) and ligated to Sall-PvuI fragment containing the Moloney leukemia virus LTR (which contains the retroviral polyadenylation signal) from HyTK-CMV-IL2.  
25

Plasmid tgCMV/CD was constructed using standard techniques (Ausubel et al., *supra*, 1987) as follows: A PvuI-NheI fragment spanning the HCMV IE94 promoter from tgCMV/hygro was ligated to a synthetic DNA fragment (prepared by annealing oligonucleotides  
30 5'-CTAGCCGCCACCATGTCTGAATAACGCTTTACAAACAATTATTAACGCCCG-3' and 5'-GTAACCGGGCGTTAATAATTGTTTGTAAGCGTTATTCGACATGGTGGCGG-3'), the BstE2-AluI fragment containing the remainder of the CD coding region from pCD2 (Mullen et al., *Proc. Natl. Acad. Sci. USA* 89:33, 1992), and the Sall-PvuI fragment containing the Moloney leukemia virus LTR (which contains the retroviral polyadenylation signal) from

HyTK-CMV-IL2. The Sall site in the latter fragment was treated with T4 DNA polymerase to render the end blunt before ligation.

Plasmid tgCMV/CD-hygro was constructed using standard techniques (Ausubel et al., *supra*, 1987) as follows: The large ClaI-Sall fragment from tgCMV/CD was ligated to a ClaI-

5 NcoI fragment amplified from tgCMV/hygro by PCR using oligonucleotides  
 5'-CCCATCGATTACAAACGTAAAAAGCCTGAACTCACC GCGAC-3' and  
 5'-GCCATGTAGTGTATTGACCGATTCC-3' (the PCR product was digested with ClaI and  
 NcoI before ligation), and an NcoI-Sall fragment containing the remainder of the *hph* coding  
 region from tgCMV/hygro/LTR.

10 Plasmid tgCMV/hygro-CD was constructed using standard techniques (Ausubel et al.,  
*supra*, 1987) as follows: The large SpeI-BstE2 fragment from tgCMV/CD was ligated to a  
 SpeI-ScaI fragment containing the *hph* coding region from tgCMV/hygro/LTR, and a synthetic  
 DNA fragment (prepared by annealing oligonucleotides  
 5'-ACTCTCGAATAACGCTTTACAAACAATTATTAACGCCCG-3' and  
 15 5'-GTAACCGGGCGTTAATAATTGTTTGTAAGCGTTATTCGAGAGT-3').

Plasmid tgCMV/CD-neo was constructed using standard techniques (Ausubel et al.,  
*supra*, 1987) as follows: The large ClaI-Asp718 fragment from tgCMV/CD was ligated to a  
 synthetic DNA fragment (prepared by annealing oligonucleotides  
 5'-CGATTACAAACGTATTGAACAAGATGGATTGCACGCAGGTTCTCC-3' and  
 20 5'-GGCCGGAGAACCTGCGTGCAATCCATCTTGTTC AATACGTTTGTAAAT-3'), and an  
 EagI-Asp718 fragment containing the remainder of the *neo* gene coding region from  
 tgCMV/neo.

Plasmid tgCMV/neo-CD was constructed using standard techniques (Ausubel et al.,  
*supra*, 1987) as follows: The large SphI-Sall fragment from tgCMV/neo was ligated to a ClaI-  
 25 NcoI fragment amplified from tgCMV/neo by PCR using oligonucleotides 5'-  
 CGAACTGTTCCGCCAGGCTC-3' and  
 5'-CCCGGTAACCGGGCGTTAATAATTGTTTGTAAGCGTTATTCGAGAA  
 GAACTCGTCAAGAAGGC-3' (the PCR product was digested with SphI and BstE2 before  
 ligation), and a BstE2-Sall fragment containing the remainder of the CD gene coding region  
 30 from tgCMV/CD.

#### B. Dominant Positive Selection of Cells containing CD Fusion Genes.

To demonstrate that the CD fusion gene encodes both neo and *hph* activities, the  
 frequencies with which the various plasmids conferred drug resistance in NIH/3T3 cells were  
 35 determined.

First, NIH/3T3 cells were grown in Dulbecco Modified Eagle Medium (DMEM; available from Gibco Laboratories) supplemented with 10% bovine calf serum (Hyclone), 2 mM L-glutamine, 50 U/ml penicillin, and 50 µg/ml streptomycin at 37°C in a humidified atmosphere supplemented with 10% CO<sub>2</sub>. For transfection, exponentially growing cells were harvested by trypsinization, washed free of serum, and resuspended in DMEM at a concentration of 10<sup>7</sup> cells/ml. Plasmid DNA (5µg) was added to 800 µl of cell suspension (8 x 10<sup>6</sup> cells), and the mixture was subjected to electroporation using the Biorad Gene Pulser and Capacitance Extender (200-300 V, 960 µF, 0.4 cm electrode gap, at ambient temperature).

Following electroporation, the cells were returned to 10 cm dishes and grown in non-selective medium. After 24 hours, the cells were trypsinized, seeded at 6 x 10<sup>5</sup> cells/10 cm dish, and allowed to attach overnight. The non-selective medium was replaced with selective medium (containing 500 U/ml of Hm or 800 µg/ml of G-418), and selection was continued for 10-14 days. The plates were then fixed with methanol, stained with methylene blue and colonies were counted. The number of colonies reported in Table 1 is the average number of colonies per 10 cm dish.

Untransfected cells were not hygromycin resistant (Hm<sup>r</sup>) or G-418 resistant (G-418<sup>r</sup>). The results indicate that the hygro-CD and CD-hygro fusion genes encode Hm<sup>r</sup>, but the activity of the CD-hygro fusion gene is lower than that of the hygro-CD fusion gene. The CD-neo fusion gene confers G-418<sup>r</sup>, but the neo-CD fusion gene does not.

Table 1

Dominant Positive Selection

Transfected	No. Hm <sup>r</sup> Colonies		No. G-418 <sup>r</sup> Colonies	
	<u>Trial 1</u>	<u>Trial 2</u>	<u>Trial 1</u>	<u>Trial 2</u>
Plasmid				
None	0	0	0	0
tgCMV/hygro/LTR	89	34	nt	nt
tgCMV/hygro-CD	96	34	nt	nt
tgCMV/CD-hygro	7 <sup>b</sup>	13 <sup>b</sup>	nt	nt
tgCMV/neo	nt	nt	28	73
tgCMV/neo-CD	nt	nt	0	0
tgCMV/CD-neo	nt	nt	29	64

nt = not tested

b = small, slowly growing colonies



The concentration of protein in the extract was determined using a Biorad protein assay kit. A 25  $\mu$ l aliquot of cell extract (or an equivalent amount of protein in a volume of 25  $\mu$ l) was then mixed with 1  $\mu$ l of [ $^{14}$ C]-cytosine (0.6 mCi/ml, 53.4 mCi/mmol; Sigma Chemical Co.), and the reaction allowed to proceed at 37°C for 1-4 h. One half of the reaction was then applied to a thin-layer chromatogram and chromatographed in a mixture of 86% 1-butanol and 14% water. Following development, the thin-layer chromatogram was exposed to Kodak X-OMAT AR X-ray film for 8-14 h. The result is shown in Figure 2.

### Example 2

### A. Construction of Retroviral Vectors.

The retroviral plasmids tgLS(+)*neo* and tgLS(+)*CD-neo* consist of the following elements: the 5' LTR and sequences through the PstI site at nucleotide 984 of MoMSV (Van Beveren et al., *Cell* 27:97, 1981); sequences from the PstI site at nucleotide 563 to nucleotide 1040 of MoMLV (Shinnick et al., *Nature* 293:543, 1981); a fragment from tgCMV/*neo* or tgCMV/*CD-neo*, containing the *neo* or *CD-neo* coding regions, respectively; sequences from nucleotide 7764 and through the 3' LTR of MoMLV (Shinnick et al., *supra*, 1981); the NruI-

AlwNI fragment from pML2d (Lusky and Botchan, *supra*, 1981), containing the bacterial replication origin; the AlwNI-AatII fragment from pGEM1 (Promega Corp.), containing the  $\beta$ -lactamase gene.

Plasmid tgLS(+)*neo* was constructed using standard techniques (Ausubel et al., *supra*, 1987) as follows: Plasmid tgLS(+)*hygro* was constructed first, by ligating an EcoRI-ClaI fragment from tgLS(+)*HyTK* to an EcoRI-Asp718 fragment from tgCMV/*hygro*, and a synthetic DNA fragment (prepared by annealing oligonucleotides 5'-GTACAAGCTTGGATCCCTCGAGAT-3' and 5'-CGATCTCGAGGGATCCAAGCTT-3'). Plasmid tgLS(+)*neo* was then constructed by replacing the NheI-HindIII fragment spanning the *hygro* gene with a NheI-HindIII fragment amplified from pSV2*neo* (Southern and Berg, *J. Mol. Appl. Gen.* 1:327, 1982) by PCR using oligonucleotides 5'-CCCGCTAGCCGCCGCCACCATGGGATCGGCCATTGAACAAGATGGATTGCAC-3' and 5'-CCCAAGCTTCCCGCTCAGAAGAAGCTCGTC-3' (the PCR product was digested with NheI and HindIII before ligation).

Plasmid tgLS(+)*CD-neo* was constructed using standard techniques (Ausubel et al., *supra*, 1987) as follows: The NheI-SalI fragment spanning the HCMV IE94 promoter and human IL-2 cDNA from HyTK-CMV-IL2 was replaced with the NheI-SalI fragment from tgCMV/*CD-neo*.

Figure 3 shows the proviral structures of the retroviral vectors tgLS(+)*neo* and tgLS(+)*CD-neo*. In the figure "LTR" signifies the long terminal repeat segments of the retroviral vector, "neo" signifies the bacterial neomycin phosphotransferase gene, and "CD-neo" represents the CD/neomycin phosphotransferase fusion gene. The *neo* and *CD-neo* genes are operably linked to the LTR transcriptional control region. The arrows show the direction of transcription from the transcriptional control regions. "A<sup>+</sup>" represents the polyadenylation sequence.

#### B. Generation of Stable Cell Lines Infected With Retroviral Vectors.

To derive stable NIH/3T3 cell lines infected with tgLS(+)*neo* and tgLS(+)*CD-neo*, the retroviral plasmid DNAs were transfected into  $\Psi$ 2 ecotropic packaging cells. The transfected  $\Psi$ 2 cells were then transferred to a 10 cm tissue culture dish containing 10 ml of complete growth medium supplemented with 10 mM sodium butyrate (Sigma Chemical Co.) and allowed to attach overnight. After 15 h, the medium was removed and replaced with fresh medium. After a further 24 hours, the medium containing transiently produced ecotropic virus particles was harvested, centrifuged at 2000 rpm for 10 minutes and used to infect NIH/3T3 cells.

Exponentially dividing NIH/3T3 cells were harvested by trypsinization and seeded at a density of  $2.5 \times 10^4$  cells/35 mm well in two 6-well tissue culture trays. On the following day, the medium was replaced with serial dilutions of virus-containing, cell-free supernatant (1 ml/well) in medium supplemented with 4  $\mu$ g/ml Polybrene hexadimethrine bromide (Sigma Chemical Co.). Infection was allowed to proceed overnight. Then the supernatant was replaced with complete growth medium. After a further 8-24 hours of growth, the infected NIH/3T3 cells were selected for drug resistance to G-418 (Gibco) at a final concentration of 800  $\mu$ g/ml ( $\text{Hm}^r$  cells). After a total of 12-14 days of growth, one tray of cultured G-418<sup>r</sup> resistant cells was fixed with 100% methanol and stained with methylene blue. The colonies were counted and the number of colonies in each well was used to establish the titers of the retrovirus present in the transiently infected supernatant (Table 2).

**Table 2**  
**Titers of Ecotropic Retroviruses Produced Transiently**  
**in  $\Psi$ 2 Packaging Cells on NIH/3T3 Cells**

<u>Virus</u>	<u>G-418<sup>r</sup></u> <u>CFU/ml</u>
tgLS(+) <sub>neo</sub>	$5 \times 10^5$
tgLS(+) <sub>CD-neo</sub>	$1 \times 10^5$

From the other tray of G-418<sup>r</sup> cells, the colonies of G-418<sup>r</sup> cells were pooled and expanded into bulk cultures for analysis. Extracts were prepared from the bulk cultures and assayed for CD activity by measuring the conversion of cytosine to uracil generally as previously described (Mullen et al., 1992), except that [<sup>14</sup>C]-cytosine was used in place of [<sup>3</sup>H]-cytosine. A 10 cm dish was seeded with  $1 \times 10^6$  cells, and the cells were incubated for 2 days. The cells were then washed in Tris buffer (100 mM Tris, pH 7.8, 1 mM EDTA, 1 mM dithiothreitol) and scraped from the dish in 1 ml of Tris buffer.

The cells were then centrifuged for 10 seconds at 14,000 rpm in an Eppendorf microfuge, resuspended in 100  $\mu$ l of Tris buffer and subjected to five cycles of rapid freezing and thawing. Following centrifugation for 5 min at 6,000 rpm in an Eppendorf microfuge, the supernatant was transferred to a clean tube. The concentration of protein in the extract was determined using a Biorad protein assay kit. A 25  $\mu$ l aliquot of cell extract (or an equivalent amount of protein in a volume of 25  $\mu$ l) was then mixed with 1 ml of [<sup>14</sup>C]-cytosine (0.6 mCi/ml, 53.4 mCi/mmol; Sigma Chemical Co.), and the reaction was allowed to proceed at 37° for 1-4 hours. One half of the reaction mixture was then applied to a thin-layer

chromatogram, and chromatographed in a mixture of 86% 1-butanol and 14% water.

Following development, the thin-layer chromatogram was exposed to Kodak X-OMAT AR X-ray film for 8-14 hours. The results shown in Figure 4 indicate that cells infected with the tgLS(+ )CD-neo retroviral vector express high levels of cytosine deaminase activity.

5

C. Negative Selection of Cells Containing the CD-neo Selectable Fusion Gene. To investigate the utility of the neo and CD-neo selectable fusion genes for negative selection, the colonies resulting from each transfection were pooled and expanded into cell lines for further analysis. The NIH/3T3 cells, or NIH/3T3 cells infected with the tgLS(+ )neo or tgLS(+ )CD-neo retroviruses were assayed for 5-FC<sup>S</sup> using a long-term proliferation assay.

10

First,  $1 \times 10^4$  cells were seeded into 10 cm tissue culture dishes in complete growth medium and allowed to attach for 4 hours. The medium was then supplemented with various concentrations of G-418 and/or 5-FC (Sigma), after which the cells were incubated for a further 10-14 days. The medium was replaced every 2-4 days. The cells were then fixed in situ with 100% methanol and stained with methylene blue.

15

Photographs of representative stained plates are shown in Figure 5. Plate a had NIH/3T3 cells grown in drug-free medium. Plate b had NIH/3T3 cells grown in medium containing 800  $\mu$ g/ml G-418. Plate c had NIH/3T3 cells grown in medium containing 100  $\mu$ g/ml 5-FC. Plate d had NIH/3T3 cells infected with tgLS(+ )neo and grown in medium containing 800  $\mu$ g/ml G-418. Plate e had NIH/3T3 cells infected with tgLS(+ )neo and grown in medium containing 800  $\mu$ g/ml G-418 and 100  $\mu$ g/ml 5-FC. Plate f had NIH/3T3 cells infected with tgLS(+ )CD-neo and grown in medium containing 800  $\mu$ g/ml G-418. Plate g had NIH/3T3 cells infected with tgLS(+ )CD-neo and grown in medium containing 800  $\mu$ g/ml G-418 and 100  $\mu$ g/ml 5-FC.

20

These results indicate that 1) uninfected NIH/3T3 cells are sensitive to G-418 and resistant to 5-FC, 2) NIH/3T3 cells infected with tgLS(+ )neo are resistant to both G-418 and 5-FC, and 3) NIH/3T3 cells infected with tgLS(+ )CD-neo are resistant to G-418 but sensitive to 5-FC.

25

## CLAIMS

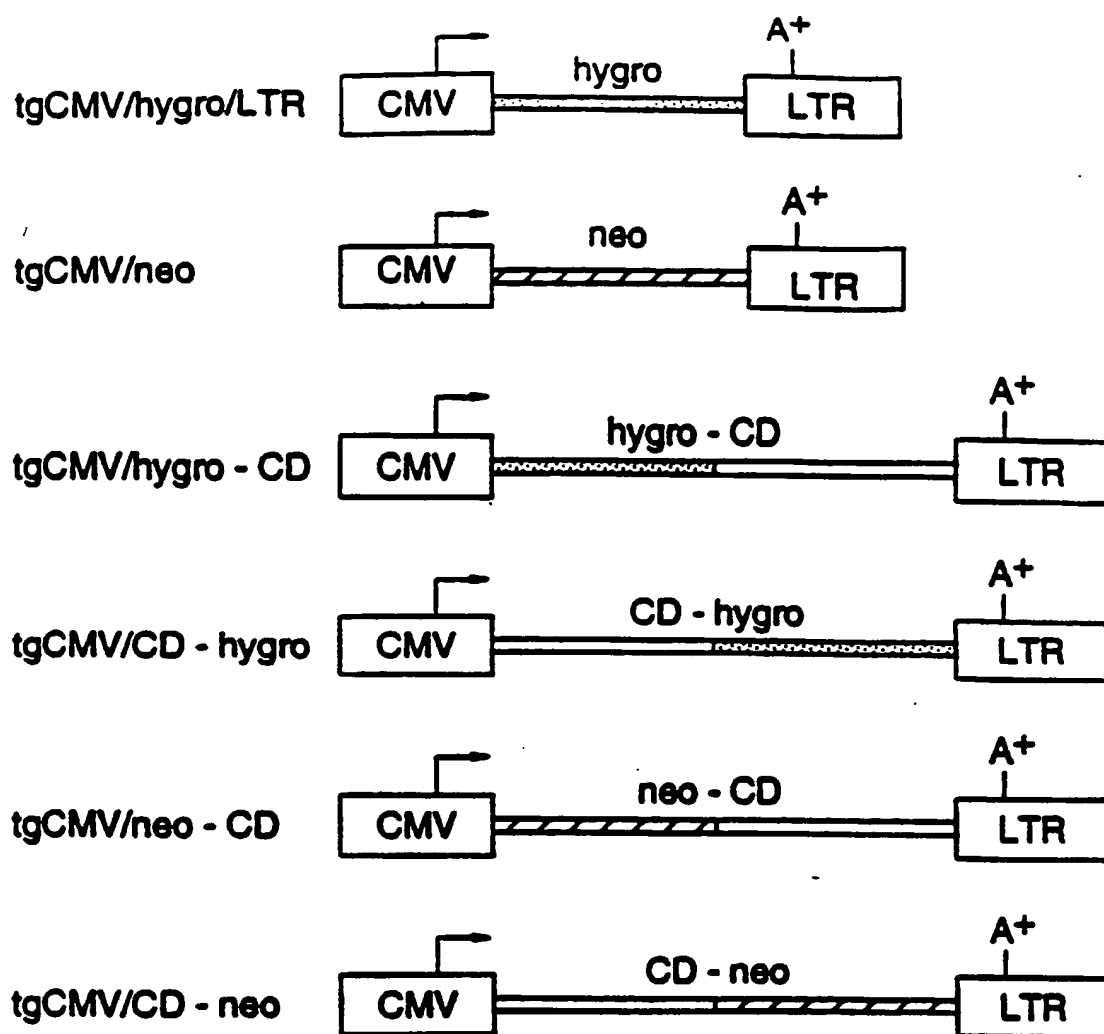
We claim:

1. A selectable fusion gene comprising a dominant positive selectable gene fused to and in reading frame with a negative selectable gene, wherein the selectable fusion gene  
5 encodes a single bifunctional fusion protein which when expressed confers a dominant positive selectable phenotype and a negative selectable phenotype on a cellular host;  
wherein the negative selectable gene is cytosine deaminase (CD).
2. A selectable fusion gene according to claim 1, wherein the dominant positive  
10 selectable gene is selected from the group consisting of *hph* and *neo* genes.
3. A selectable fusion gene according to claim 2, wherein the dominant positive selectable gene is *neo*.
- 15 4. A selectable fusion gene according to claim 3 encoding the sequence of amino acids 2-690 of SEQ ID NO:2.
5. A selectable fusion gene according to claim 3 encoding the sequence of nucleotides 4-2073 of SEQ ID NO:1.  
20
6. A recombinant expression vector comprising a selectable fusion gene according to claim 2.
7. A recombinant expression vector comprising a selectable fusion gene according to claim 3.  
25
8. A recombinant expression vector comprising a selectable fusion gene according to claim 4.
- 30 9. A recombinant expression vector according to claim 6, wherein the vector is a retrovirus.
10. A recombinant expression vector according to claim 7, wherein the vector is a retrovirus.

-27-

11. A recombinant expression vector according to claim 8, wherein the vector is a retrovirus.
12. A cell transduced with a recombinant expression vector according to claim 6.
13. A cell transduced with a recombinant expression vector according to claim 9.
14. A method for conferring a dominant positive and negative selectable phenotype on a cell, comprising the step of transducing the cell with a recombinant expression vector according to claim 6.
15. A method for conferring a dominant positive and negative selectable phenotype on a cell, comprising the step of transducing the cell with a recombinant expression vector according to claim 9.

1/6



**CMV = Cytomegalovirus**

**hygro = hygromycin transferase gene**

**neo = bacterial neomycin phosphotransferase**

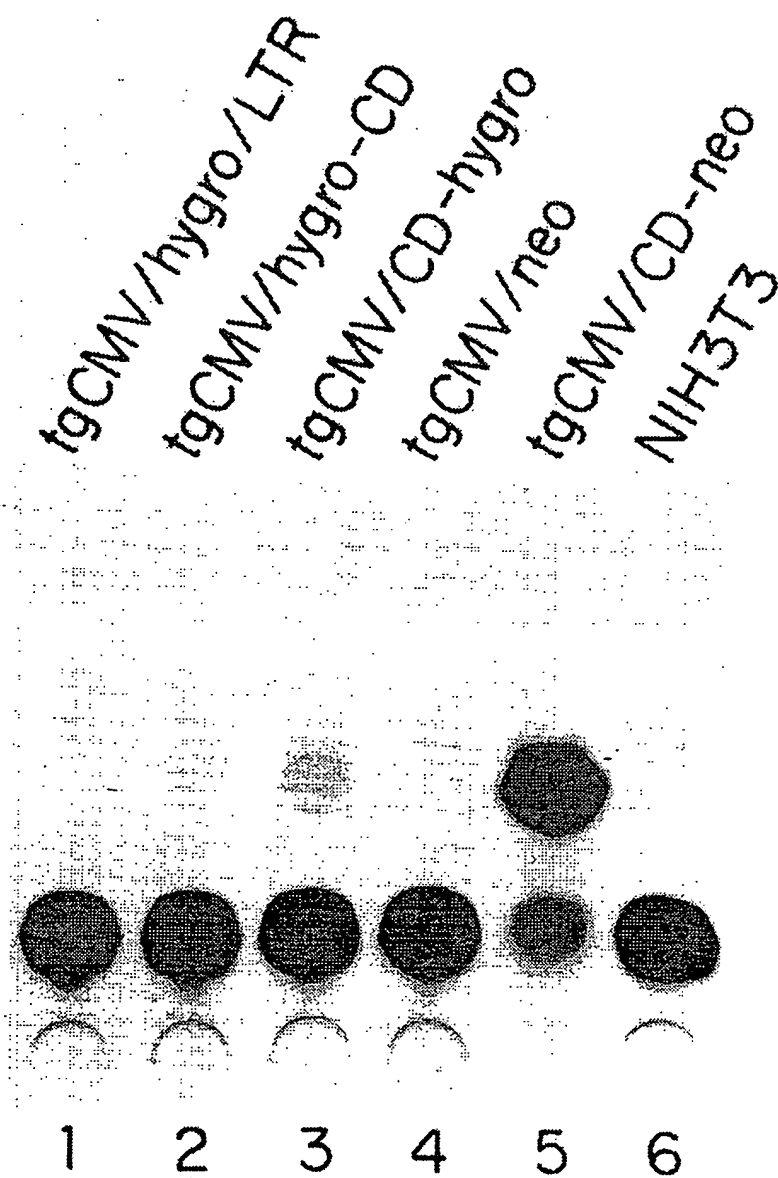
**CD = cytosine deaminase gene**

**LTR = Long Terminal Repeat**

**A+ = polyadenylation signal**

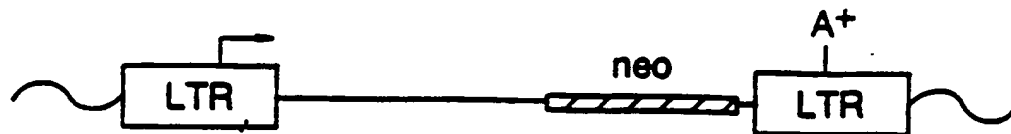
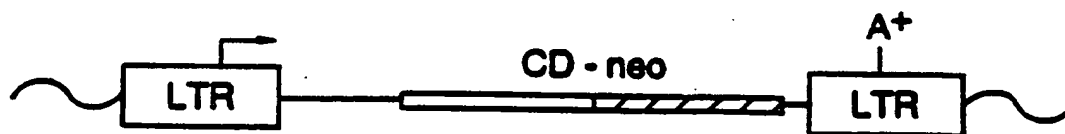
**FIGURE 1**

2/6

**FIG. 2**



3/6

tgLS (+) neotgLS (+) CD-neo

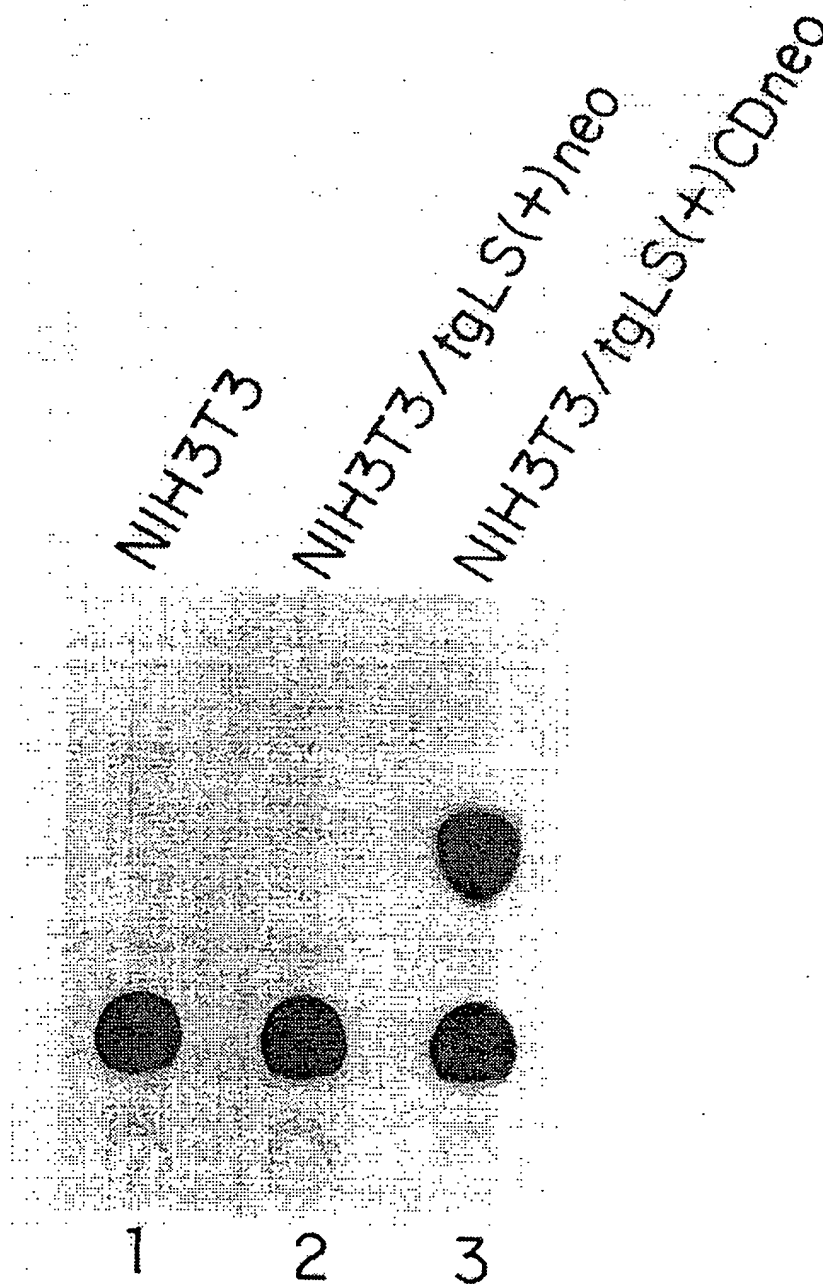
**LTR = Long Terminal Repeat**

**neo = bacterial neomycin phosphotransferase**

**CD = cytosine deaminase gene**

**A+ = polyadenylation signal**

**FIGURE 3**



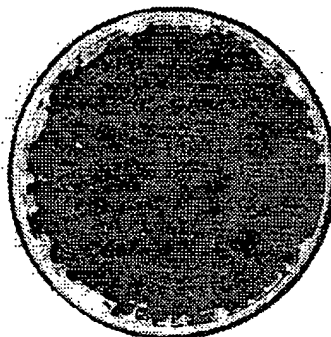
**FIG. 4**

SUBSTITUTE SHEET (RULE 26)

5/6

**FIG. 5a**

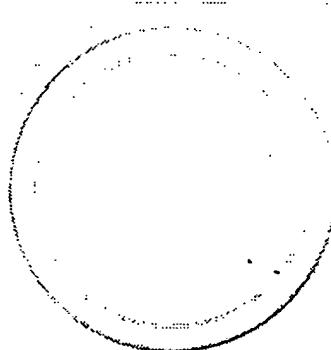
Uninfected:



No Drug

**FIG. 5b**

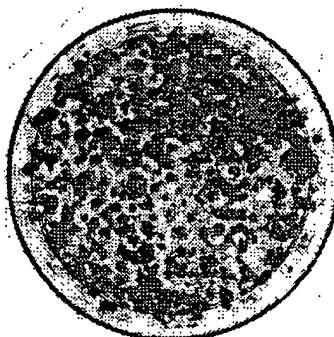
Uninfected:



G-418

**FIG. 5c**

Uninfected:



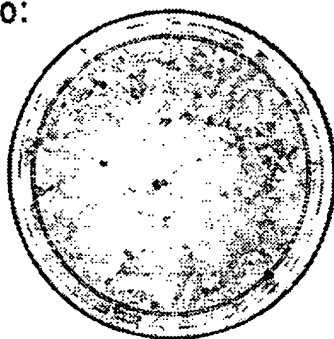
5-FC

SUBSTITUTE SHEET (RULE 26)

THIS PAGE BLANK (119PT0)

6/6

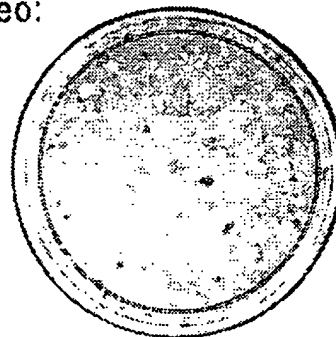
tgLS(+)  
neo:



G-418

**FIG. 5d**

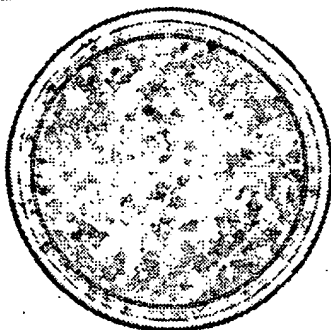
tgLS(+)  
neo:



G-418 + 5FC

**FIG. 5e**

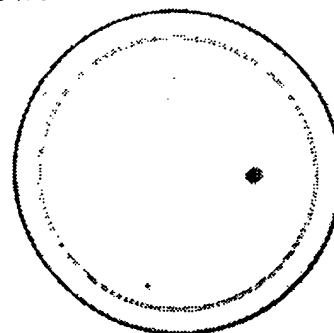
tgLS(+)  
CDneo:



G-418

**FIG. 5f**

tgLS(+)  
CDneo:



G-418 + 5FC

**FIG. 5g**